Legal Information Management, 16 (2016), pp. 219–224
© The Author(s) 2016. Published by British and Irish Association of Law Librarians
doi:10.1017/S1472669616000499

Big Data Technology, Evolving Knowledge Skills and Emerging Roles

Abstract: In a world where articles and tweets are discussing how artificial intelligence technology will replace humans, including lawyers and their support functions in firms, it can be hard to understand what the future holds. This article, written by Alex Smith, is based on his presentation at the British and Irish Association of Law Librarians conference in Dublin 2016 and looks at demystifying the emerging technology boom and identifies the expertise needed to make these tools work and be deployed in law firms. The article then looks at the skills and expertise of the knowledge and information teams, based in law firms, and suggests how they are ideally placed to lead these challenges as a result of their domain expertise and their existing, well defined skills that are essential to this new generation of technology. The article looks at the new technical environment, the emerging areas of products and legal problems, the skills needed for the new roles that this revolution is creating and how this could fit into a reimagined knowledge team.
Keywords: big data; information technology; knowledge management; law firms

INTRODUCTION

The last two years has seen unprecedented hype around artificial intelligence (AI), big data, machine learning and other computing buzzwords within the legal press, on social media and at every conference on the circuit. Combined with a transformation within legal IT as a function (within law firms and businesses) with the emergence of cloud computing and with methodologies to develop, like Agile and Lean, it feels like the time is right for many of the predictions about computers and robots to come true where legal tasks, automating processes and solving society's issues of interacting with the law are concerned. Whilst many of the techniques being talked about (like natural language processing) have been around for over 30 years (and there is a natural inclination to worry about another hype cycle) it feels like people have the tools and inclination to transform how law is delivered through computing. We are seeing daily on Twitter a brash new breed of start-ups, thinkers, progressive lawyers, academics (on both sides of the Atlantic), technology companies and data scientists talking openly about how their idea is going to change some component of the way law firms operate, or deliver legal services, or indeed replace lawyers and firms. So, if this is finally 'the moment', where does this leave the professionals who have made law firms work until now? What does this mean for the lawyers, for the support lawyers, for the knowledge teams and information specialists?

This article is a synthesis of three years of talking at the British and Irish Association of Law Librarians conference on the topic of what big data and associated technology may do to law, to the way lawyers work, the way firms operate and, importantly, what it means for the teams of knowledge managers and information professionals. Over the course of the three years the landscape has changed, both in the public realm of the legal press and behind the scenes, in relation to start-ups, emerging tech companies, the large information companies and the firms themselves. Whether people are publicising their efforts, or things are just 'happening' behind closed walls, there is a lot more knowledge right now about the art of 'the possible', than there was three years ago (since I first stood up for a 10 minute lightening session at the BIALL conference in Harrogate). At LexisNexis three years ago I, and a brilliant US colleague (on our technology architecture team), had only just started playing with linked data on LexisNexis Australia's legal corpus that would start to change how I thought about data and products. I was only just starting to frame a project on a UK legal taxonomy that would take me to the highs and (many) lows of understanding of machine learning techniques. I had only just begun to talk to John Sheridan and his National Archives team on the 'Big Data for Law' project and how he was thinking about connecting and understanding the legislative books through connections and patterns. It seems a long time ago but that was probably because of the pace of understanding and the nature of the learning curve that we all have been on.

THE LEARNING CURVE

I will start by summarising what I stood up to talk about on a rainy Saturday morning in Harrogate in 2014. Having heard John Sheridan speak a few months before at Lexis House on the future of legislation, and using big data tools

https://doi.org/10.1017/S1472669616000499 Published online by Cambridge University Press

to analyse patterns and trends in legislation, I borrowed (with credit) John's analogy of the Hubble Space Telescope. The Hubble telescope was sent into space in 1990 to see what the Earth's atmosphere prevents land bound telescopes from seeing. It was a great piece of engineering but no-one knew what it would find, almost 30 years later the results have been staggering. What is the analogy? Often, in our desire for certainty we want to know what we will get (and when) from any technology project so we never build the Hubble in the legal context … imagine the business case conversation, "I want to build something that may or may not find anything, let's just do it and see what happens". What is interesting is that, with the emergence of big data technology, we are now at that point, especially with cloud computing and the elasticity of that cloud technology. Under the (rather) open question of 'what is the future' I went for the bold approach of not discussing what I knew was in the LexisNexis product roadmap, but instead what I thought would be happening in 2025.

By trying to focus on four areas to make this real, these now feels like the building blocks of what is called artificial intelligence, namely (a) finding and classifying information on a large scale (b) finding 'things' like people, companies, judges, financial figures like court awards, etc in unstructured data (c) finding best practice in working practice like deals and matters and (d) expert systems based on learning from what lawyers do in systems. At the same time, the volume of data available is increasing deal by deal or social microblog at a time. The problem is that the environment of data is changing on a daily basis both from public sources (like the ever increasing flow of case law) and within the law firms' ecosystems of deals and advice. The answer will increasingly be across open data, published information and within the firm/in-house ecosystem.

So to move forward a year and to the BIALL Conference 2015, held in Brighton, for another ten minute slot, and with a year's more experience in the field with that brilliant LexisNexis technology architect, we were getting back to what the internet started as … Linked Data. Instead of relying on search, we asked, what if legal knowledge could be linked together and see (a)

What will have already happened in 2025?	What will big data for law be doing in 2025?
Search will find answers not keywords	Finding connections
Legal processes will be mapped and drive efficiency	Finding trends in data
Legal project management will be the norm	Finding new data/views on existing data
What can be automated will have been	Comparing and contrasting data
Cloud will not be a discussion point, it will just be where things are hosted	Benchmarking best practice
	Finding user behaviour and patterns

Figure 1: 2025 and big data.

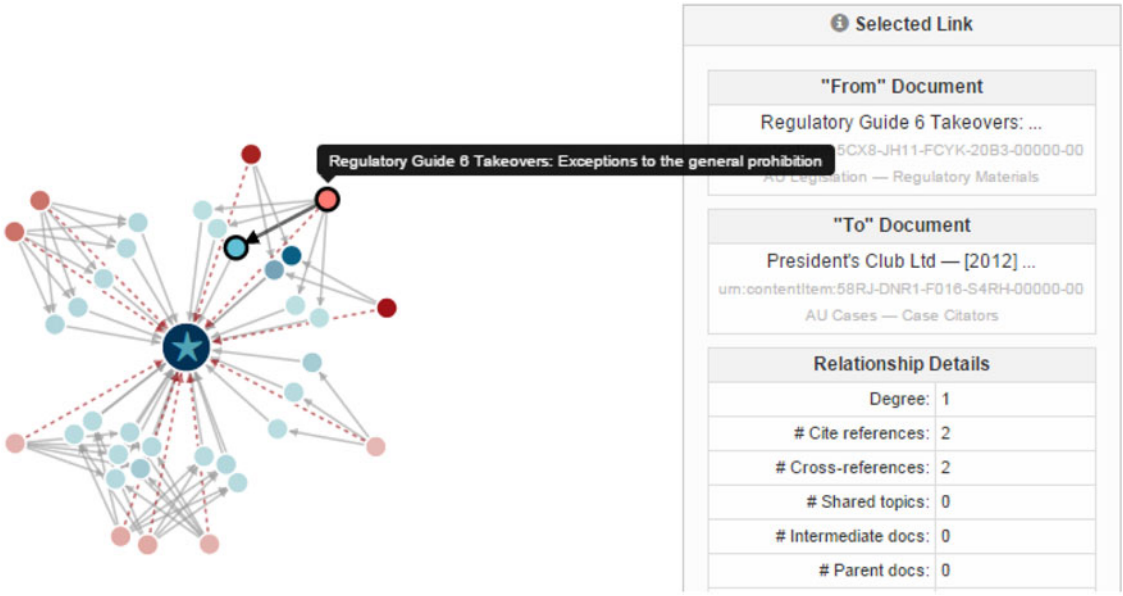


Figure 2: Example screenshot from the demo showing how a regulatory code is related to a case.

https://doi.org/10.1017/S1472669616000499 Published online by Cambridge University Press

how it connected and (b) see patterns in the data – both, how you would expect it be connected and opening up for serendipity in what could be found. I went through a live demo of a graph database of 6 million Australian legal documents connected together that generated over 900,000,000 connections between the content – connections between legislation, caselaw, commentary and practical guidance. We were starting to connect the 'what the law is about to' to the 'how to use the law'. What was interesting about the feedback I received after the session was how powerful the data visualisation was for an audience of knowledge and information professionals.

BECOMING REAL IN 2016

From 2015 to 2016 it feels that the legal industry has really embraced next generation technology and, within that, it has started to explore big data and artificial intelligence but without always explaining what is meant by those terms; and more importantly what that means to people who have traditionally dealt with legal information, both at the sharp end as a lawyer trying to find the 'answer' at 11pm in the office or those organising the information in the knowledge teams. Start-ups seem to be pivoting what their businesses do at a sharp pace, one day a company is answering legal questions and the next they are offering more targeted research, one day a company is a M&A deal platform and the next they are an automation platform. How do you, as a knowledge and information professional, know what is coming next that could take your job? What will never work? What will do that particular task that you said should have been done by computers a decade ago, but you have had to do by hand for those 20 years? The trend is not something people can ignore, there are real products and technologies to embrace. Some of them even come from other industries, like blockchain and smart contracts from Fin Tech. Those industries are often the firm's sector clients, so understanding this technology is a must. With the need to provide value-add services to the firm's clients, both written knowledge and knowledge embedded in these technologies is essential to the firm-client relationship.

The other trend in the last year has been the rise of the new role titles in firms and alternative business structures. Knowledge engineers, process architects, legal project managers and innovation managers are appearing everywhere. Job titles with a combination of technology DNA (like architects and engineers), lean six sigma DNA (like process or efficiency) and project management (continuous improvement), in addition to innovation role titles, are coming up thick and fast. Much of this has come from America with several universities now teaching law and technology like Michigan State University (MSU); and these graduates are now going into practice. For firms like Seyfarth Shaw, lean experiment is now just the normal way to do things at that firm. So again, the question arises, what does this all mean for the knowledge and information profession and their role in this changing world?

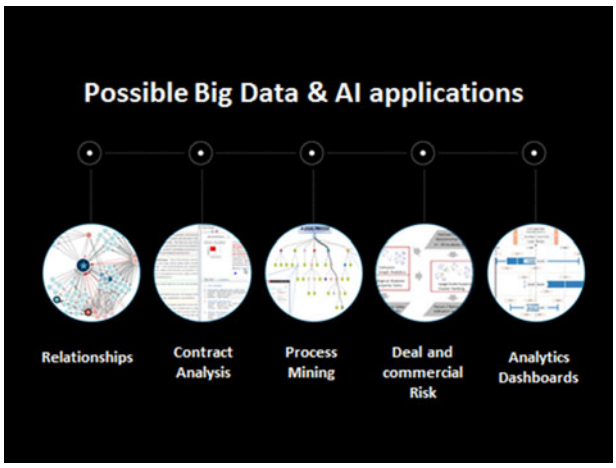


Figure 3: Possible big data and AI applications.

WHAT DOES THE WORLD LOOK LIKE TODAY? EMERGING THEMES AND REAL EXAMPLES

Big data applications are all around us today from the tiles on your google results pointing to a film star's Wikipedia page to the recommendation engine on Amazon that means you end up buying one too many boxset DVDs; to your Facebook feed that is curating itself based on who you interact with and whose posts you like; to the point that a 'friend' can drop almost completely out of your feed over time. But the question that keeps coming back is, how is this relevant to law. A starting point is to deconstruct what people are talking about under the heading of 'artificial intelligence' and then think of how to apply that to lawyers who are the end users of the possible applications. Here is a simple breakdown of what is in the AI bucket at present:

1. Big data: the ability to deal with lots of data. Sounds simple but we are only finally hitting the moment when this is viable for applications; it is going to be the foundation of everything else.
2. Analytics: measuring and understanding trends in data and giving that to lawyers to make data-driven decisions or support their human decisions.
3. Machine Learning: moving from writing rules to find and classify, allowing the machines to learn from data the patterns and clusters.
4. Natural Language Processing: vital to a world in which lawyers do most things using words and language; the ability to understand language constructions. This could include speech recognition as well as textual understanding.
5. Data visualisation: the ability to visually display and interact with data for analysis and communication
6. Decision logic: encoding the law into a series of logical outcomes; initially this already sits behind document automation but it is emerging in the efficiency world and it could go further.

https://doi.org/10.1017/S1472669616000499 Published online by Cambridge University Press



Figure 4: A map of the areas of applications for lawyers.

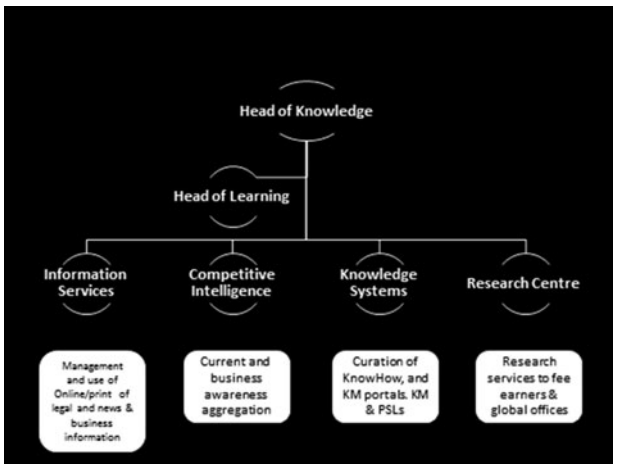


Figure 5: The traditional knowledge team.

What is interesting is that some, or all, of these could be deployed to meet legal or client problems, but who knows what to deploy for what purpose? This is the new toolkit for the savvy knowledge professional … I will discuss later some of the aspects underlying the traditional information architecture and the understanding principles that sit beneath this technology.

It is essential to understand how all this technology could be applied to the working lives of lawyers before spending money on new tech. The last 12 months has finally seen the beginning of real applications of this technology both in the UK and the US. The emerging areas are:

1. **'deal point' analytics**: analysis of contractual or diligence data – in forms and documents – for key data points, extracting key data points from semi-structured or unstructured contract data.
2. **case analytics**: analysis of parts of case law and case history and predict or signpost outcomes or length of cases.
3. **contract analysis**: diligence tools, contract drafting and lifecycle tools that are starting to emerge, unlocking data within legal contracts.
4. **advice and answers**: lawyers are asking search engines questions and want answers that help give something near to 'advice'.
5. **process**: codifying law into decision-based knowledge bases that drives legal processes.
6. **visualisation**: making content and experiences more dynamic and visual for the visual people amongst us.

Again, none of these are mutually exclusive. Many of the emerging products on both sides of the Atlantic are using combinations of these techniques – for example, in the US, Lex Machina delivers case analytics but delivered through a high visual interface, and in the UK RAVN is applying linked data visualisation and contract analysis on top of enterprise search.

THE IMPACT ON THE KNOWLEDGE TEAM

With the challenges to create these applications for lawyers, or for their clients, what is the impact of this and the emergence of new role titles on the knowledge function. Traditionally, this function has covered (a) information services (b) research specialism (c) building knowledge systems and (d) increasing 'aggregating focused updates' and business intelligence.

Depending on the firm and their challenges, a combination of new roles are emerging to take on the new challenges of efficiency, client services and a new approach to technology and IT. So what are these roles? Let us take an imaginary law firm taking on some of these roles. We will look at their role and examples of what they are currently working on.

PROCESS ENGINEER:

Role: captures legal processes within the firm using business process mapping techniques, improves processes using previous matter metrics, works to turn explicit knowledge into decision logic technology and runs process mapping sessions with key clients.

Current projects: working with the corporate practice area KM & PSL teams to map key repeatable, high value transactions, analysing the last years' transactions to see matter cost and improvement areas and working with a key bank client to map 3 core transactions into a self-serve portal.

LEGAL KNOWLEDGE ENGINEER:

Role: designing and building decision support applications, overseeing development and maintenance of software tools built in-house and working directly with clients and attorneys to implement solutions that solve their operations challenges.

https://doi.org/10.1017/S1472669616000499 Published online by Cambridge University Press

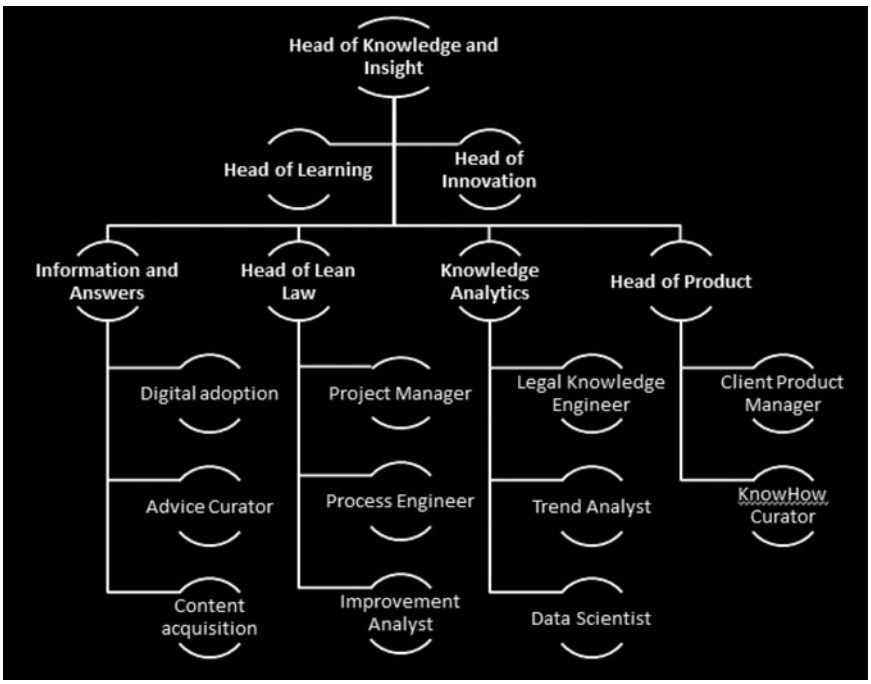


Figure 6: The knowledge team of the future.

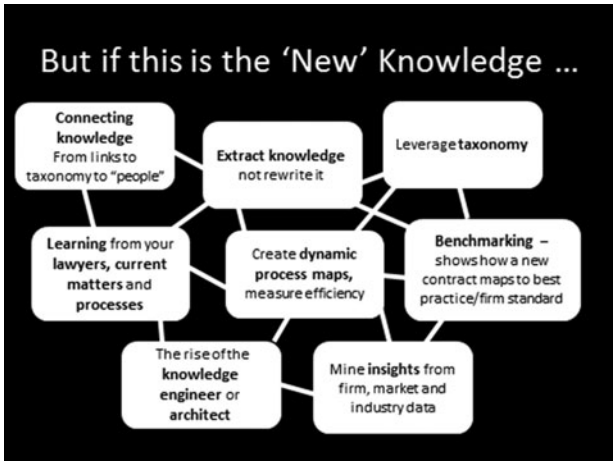


Figure 7: 'New' Knowledge…

Current projects: leading projects to get key repeatable matters into a legal task management tool, consulting on and developing custom contract management tools for key clients (using AI tools) and working on a major client's repeatable sales contract generation tool and management reporting tool.

DATA SCIENTIST:

Role: examines the data environment of the law firm, examines where the data is that the firm needs, models data across multiple systems and platforms, works with partners, knowledge management, business development and finance teams to help data driven decisions and looks at external data sources for Big Data/AI projects (eg learning sets, taxonomy).

Current projects: leading data project around matter profitability – trying to find financial metrics and matter type info (taxonomies), working for knowledge team on the way to generate explicit knowledge from previous research enquiries and junior lawyer research memos.

AGILE PRODUCT MANAGER:

Role: manages internal systems development in a user centric and agile manner, owns development roadmap for (eg) knowledge and information tools, delivers client 'value-add' products for key clients and leverages internal knowledge and external technology (including big data tools).

Current projects: working on a search project for the law firm's KM & document systems, working on relevancy and linking of results and ongoing enhancements to big data risk checking tool for major real estate client.

INFORMATION EXPERT:

Role: understands legal information, advocates good information resources big data/AI tools, understands both content and metadata and how these can help, provides training content for AI tools, checks, validate and involved in machine learning (teaching the machines).

Projects: working on a knowledge graph for the law firm in order to deliver linked answers, working on curating the firm's recommendations engine for the

https://doi.org/10.1017/S1472669616000499 Published online by Cambridge University Press

training and skills objectives of trainees, junior and associate lawyers and working on the sector taxonomy project for the firm to provide industry insight portals.

KNOWLEDGE TEAM OF THE FUTURE

So, as these roles emerge where do they sit? It is one view that these roles reside with the knowledge team as a centre of knowledge about information, the capture and re-use of knowledge assets and process re-use. What might this team look like? Looking at the diagram of a future knowledge team the department is a team for (a) information and answers (b) process and efficiency (c) analytics and (d) product management. These would, of course, be tied to the firm's strategy, so may not exist for some firms, or flex based on priorities.

COOL JOB TITLES, OLD PRINCIPLES

So, why is the knowledge team the place where this will happen and sit? It is because a lot of the new artificial intelligence and big data technology, tools and techniques require many of the skills that have been at the heart of the knowledge and information management community for decades. Why? Because AI and big data is logical, requires understanding of information architecture and a knowledge of the domain (ie. the law and lawyers). The following skills seem as relevant in this new world as they have been to date:

1. **extracting explicit from implicit knowledge**: understanding where tacit knowledge is in the law firm, re-using that knowledge to train the machines and extracting processes and logic from lawyers.
2. **learning and training sets**: machines need to learn and the knowledge team are the experts of where the good content to learn is in the law firm.
3. **logical people**: new technology and systems to capture and re-use legal logic, even creating and maintaining flowcharts and checklists out of process mapping is logic.
4. **organisation of data**: classifying and structuring systems is key to these systems learning the information architecture of the firm.
5. **understanding 'legal' English**: understands the legal language domain for law and for areas of practice in order to help the machines learn.
6. **quality and testing of the outputs of the machine learning**: the tools needing testing for the accuracy and to learn from the testing (supervised machine learning).

These skills are at the core of those skills that already exist in knowledge and information management; the ability to know what is good information, the right sources of information, classification and organisation of information, legal domain knowledge, ability to look across the domain (rather than practice area specific) and understanding how to look for new sources of information. There is plenty to learn regarding the technology, but the teams are ideally placed to be at the core of a law firm looking to leverage artificial intelligence and big data. There will be lots of opportunities to learn new skills and pivot into new roles like knowledge architects and data scientists, as well as the opportunity for working closely in agile technical teams; great for evolving career paths! There will also be new people applying for roles within the knowledge teams that will bring a different perspective on legal problem-solving and how to look at data; but at the same time these people will need to understand and navigate the legal domain and the domain of legal information. In many ways, the principles of knowledge management, of collecting, organising, using, learning and improving knowledge bases are at the core of how to train the machines to build the next generation of legal knowledge and intelligence. There is a great opportunity to redefine roles and skills within this emerging world, especially when we all realise that AI and the emerging technology is not magic! In fact it is a lot of hard work, needing a clear understanding of the legal domain and an understanding of the information architecture of knowledge that is hiding in law firms, in public data and among publishers; three things which knowledge and information teams have a lot of experience and expertise in. Here is to a brave new world … with some old world skills and expertise.

But are these challenges really 'new'?

Biography

Alex Smith is Innovation Manager at Reed Smith, working on innovation and transformation of legal services and the creation of new products, including digital channels, big data, artificial intelligence and workflow technologies. He previously worked at LexisNexis for 17 years, honing his expertise with online legal information and productivity tools. As a Senior Lead for Platform Innovation he combined customer-experience insight with developing next generation technology. By exploring the way lawyers work in an evolving legal industry, he develops products that support their collaborative, social and linked-data needs. This includes shared-working environments (the cloud), that encourage the contribution, distribution and adaptation of knowledge. Alex worked recently on the visualisation of legal and contract data on LexisNexis' Linked Data and AI tools. He is also working on the capture of legal knowledge as process and decision logic to drive efficiency and better access to justice. This talk was given when working at LexisNexis in June 2016.

https://doi.org/10.1017/S1472669616000499 Published online by Cambridge University Press